

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 09-152892

(43)Date of publication of application : 10.06.1997

(51)Int.Cl. G10L 5/04  
 G10L 3/02  
 G10L 7/02  
 G10L 9/00

(21)Application number : 08-238235

(71)Applicant : NIPPON TELEGR & TELEPH CORP  
 <NTT>

(22)Date of filing : 09.09.1996

(72)Inventor : ABE MASANOBU

(30)Priority

Priority number : 07248144 Priority date : 26.09.1995 Priority country : JP

## (54) VOICE SIGNAL DEFORMATION CONNECTION METHOD

## (57)Abstract:

PROBLEM TO BE SOLVED: To provide a voice signal deformation connection method capable of connecting the voice messages of different voice quality each other without the sense of incongruity.

SOLUTION: Two voice signals 101 and 102 obtained by making two speakers utter the same text are connection-processed. Synthetic voice 103 resulted from this processing is constituted of the voice section of the speaker 1, a deformation connection section and the voice section of the speaker 2. Even in the case that the voice quality of the two speakers is widely different, when the extent of deformation at one time is small, a listener side does not feel the sense of incongruity so much. Then, the voice is connected by repeating the deformation without the sense of incongruity stopwise for several times. That is, in the voice message obtained by connection, the voice quality is gradually changed over the prescribed time of the deformation connection section.



BEST AVAILABLE COPY

## LEGAL STATUS

[Date of request for examination] 17.05.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3465734

THIS PAGE BLANK (USPTO)

[Date of registration] 29.08.2003

[Number of appeal against examiner's decision  
of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright © 2006

**THIS PAGE BLANK (USPTO)**

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-152892

(43) 公開日 平成9年(1997)6月10日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所	
G 1 0 L	5/04		C 1 0 L	5/04	D
	3/02			3/02	D
	7/02			7/02	D
	9/00			9/00	E

審査請求 未請求 請求項の数12 O L (全 10 頁)

(21) 出願番号 特願平8-738235

(22) 出願日 平成8年(1996)9月9日

(31) 優先権主張番号 特願平7-248144

(32) 優先日 平7(1995)9月26日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000004226

日本電信電話株式会社

東京都新宿区西新宿三丁目19番2号

(72) 発明者 阿部 匡伸

東京都新宿区西新宿三丁目19番2号 日本

電信電話株式会社内

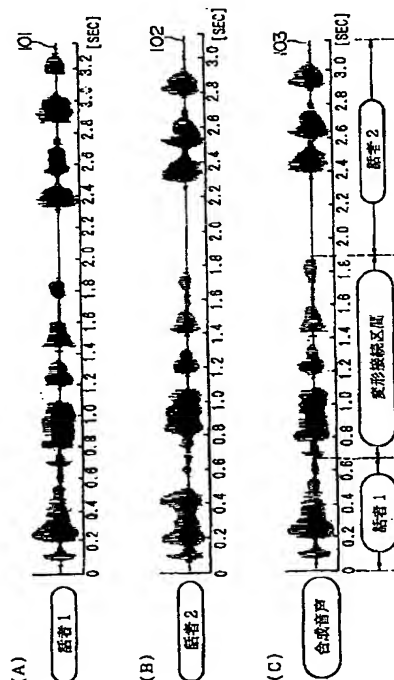
(74) 代理人 弁理士 志賀 正武

(54) 【発明の名称】 音声信号変形接続方法

(57) 【要約】

【課題】 声質の異なる音声メッセージ同士を違和感無く接続することができる音声信号変形接続方法を提供する。

【解決手段】 2人の話者に同じテキストを発声させて得られる2つの音声信号101及び102を接続処理する。この処理の結果生成された合成音声103は、話者1の音声区間、変形接続区間、話者2の音声区間とから構成される。2人の話者の声質が大きく違う場合であっても、一度に変形する度合いが小さければ、聞く側はあまり違和感を感じない。そこで、違和感を感じない程度の変形を段階的に複数回繰り返すことによって音声の接続を行う。すなわち、接続により得られる音声メッセージは、変形接続区間の所定時間に渡って徐々に声質が変化していく。



BEST AVAILABLE COPY

## 【特許請求の範囲】

【請求項1】 互いに異なる声質に属する2つの音声信号同士を接続する処理において、音声信号の特性を示すパラメータの値を、一方の音声信号の特徴を示す値から他方の音声信号の特徴を示す値へと所定時間にわたって徐々に変化させることにより、2つの音声信号を接続することを特徴とする音声信号変形接続方法。

【請求項2】 請求項1記載の音声信号変形接続方法において、前記パラメータの値を変更する所定時間にわたり、前記互いに異なる声質の話者に同一のテキストを読み上げさせ、これにより得られる2種類の音声データを用いて前記パラメータ値の変更を行うことを特徴とする音声信号変形接続方法。

【請求項3】 請求項1記載の音声信号変形接続方法において、前記互いに異なる声質の音声信号は音声合成装置による発声により得られるものであることを特徴とする音声信号変形接続方法。

【請求項4】 請求項1記載の音声信号変形接続方法において、前記互いに異なる声質の音声信号は、一方は人間による発声、もう一方は音声合成装置による発声により得られるものであることを特徴とする音声信号変形接続方法。

【請求項5】 請求項1記載の音声信号変形接続方法において、前記パラメータは音声のスペクトルであり、前記所定の時間にわたって前記スペクトルを徐々に変形させることを特徴とする音声信号変形接続方法。

【請求項6】 請求項5記載の音声信号変形接続方法において、前記音声のスペクトルの変形に関しては、前記2つの音声信号間の対応する音素内のピッチの対応を決定し、対応したピッチ毎に、ある周波数を境界周波数として、一方の音声信号のスペクトルにおける前記境界周波数より高域の部分と、他方の音声信号のスペクトルにおける前記境界周波数より低域の部分とを結合させたスペクトルを生成し、生成されたスペクトルを当該ピッチにおけるスペクトルとすると共に、前記境界周波数を単位時間毎に変化させることを特徴とする音声信号変形接続方法。

【請求項7】 請求項6記載の音声信号変形接続方法において、前記境界周波数の変化は、単位時間毎に一定量増加するように行われることを特徴とする音声信号変形接続方法。

【請求項8】 請求項6記載の音声信号変形接続方法において、

前記境界周波数の変化は、変化開始時の値から変化終了時の値まで徐々に増加するように行われ、前記変化開始時に近い相対的に低域の段階では比較的ゆっくりと、前記変化終了時に近い相対的に高域の段階では比較的早く変化させることを特徴とする音声信号変形接続方法。

【請求項9】 請求項1記載の音声信号変形接続方法において、

前記パラメータは音声の基本周波数であり、前記所定の時間にわたって前記基本周波数を徐々に変化させることを特徴とする音声信号変形接続方法。

【請求項10】 請求項9記載の音声信号変形接続方法において、

前記音声の基本周波数の変化に関しては、前記各音声信号の平均基本周波数を求め、両者の平均基本周波数の差とパラメータを変更する前記所定時間とに基づいて、単位時間当たりに変更すべき基本周波数の値を求め、

この値を変化量の単位として、一方の音声信号の平均基本周波数から他方の音声信号の平均基本周波数へと単位時間毎に変化させることを特徴とする音声信号変形接続方法。

【請求項11】 請求項1記載の音声信号変形接続方法において、

前記パラメータは、音声のスペクトルと音声の基本周波数であり、前記音声のスペクトルの変形に関しては、前記2つの音声信号間の対応する音素内のピッチの対応を決定し、

対応したピッチ毎に、ある周波数を境界周波数として、一方の音声信号のスペクトルにおける前記境界周波数より高域の部分と、他方の音声信号のスペクトルにおける前記境界周波数より低域の部分とを結合させたスペクトルを生成し、生成されたスペクトルを当該ピッチにおけるスペクトルとすると共に、

前記境界周波数を単位時間毎に変化させるようにし、

前記音声の基本周波数の変化に関しては、

前記各音声信号の平均基本周波数を求め、

両者の平均基本周波数の差とパラメータを変更する前記所定時間とに基づいて、単位時間当たりに変更すべき基本周波数の値を求め、

この値を変化量の単位として、一方の音声信号の平均基本周波数から他方の音声信号の平均基本周波数へと単位時間毎に変化させるようにすることを特徴とする音声信号変形接続方法。

【請求項12】 請求項11記載の音声信号変形接続方法において、

前記音声のスペクトルと基本周波数とを並行して変化させることを特徴とする音声信号変形接続方法。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、録音編集型の音声メッセージの作成において、音声メッセージの追加、変更を効率よく行うことを可能とし、音声メッセージを用いたシステムの構築、維持の経済化をはかることができる音声信号変形接続方法に関する。

#### 【0002】

【従来の技術】現在、駅の構内アナウンスや、道路の渋滞などの情報を知らせるハイウェイラジオや、情報検索における音声ガイダンス等のサービスには、音声メッセージが使われている。これらの音声メッセージは、予め人間が発声した音声を録音し、この音声を継ぎはぎすることによって作成されている。

【0003】係る音声メッセージの作成において、既に作成された音声メッセージとは異なる新たな音声メッセージが必要となり、その必要な音声メッセージが録音されていない場合には、新たに音声を追加録音する必要がある。この場合、既録音の音声と新規録音の音声との間で声質が急激に変化することなく自然につながるように、以前に発声した話者と同じ人に追加分を発声してもらう必要があった。

#### 【0004】

【発明が解決しようとする課題】しかしながら、同一話者であっても、以前の収録から年月が経っている等の理由で以前の声質と異なり、新旧メッセージの継ぎはぎにより聞き苦しさや予想される場合には、すべての音声メッセージを再び収録および作成し直す必要があった。また、以前に発声した人が不在の場合には、他の話者に代わりて発声してもらい、全ての音声メッセージを再び収録し直す必要があった。また、上記のような音声メッセージは音声合成装置を用いて作成することも可能であるが、この場合も、音声合成装置が異なる等の理由により互いに異なる声質となって出力された音声信号同士を接続する場合に、同様の問題が生じる。

【0005】本発明は、このような背景の下になされたもので、声質の異なる音声メッセージ同士を違和感無く接続することができ、音声メッセージの追加、変更を効率よく行うことができる音声信号変形接続方法を提供することを目的とする。

#### 【0006】

【課題を解決するための手段】上記課題を解決するために、請求項1による音声信号変形接続方法においては、互いに異なる声質に属する2つの音声信号同士を接続する処理において、音声信号の特性を示すパラメータの値を、一方の音声信号の特徴を示す値から他方の音声信号の特徴を示す値へと所定の時間にわたって徐々に変化させることにより、2つの音声信号を接続することを特徴とする。

【0007】また、請求項2による発明は、請求項1記載の音声信号変形接続方法において、前記パラメータの値を変更する所定時間にわたり、前記互いに異なる声質

の話者に同一のテキストを読み上げさせ、これにより得られる2種類の音声データを用いて前記パラメータ値の変更を行うことを特徴とする。

【0008】また、請求項3による発明は、請求項1記載の音声信号変形接続方法において、前記互いに異なる声質の音声信号は音声合成装置による発声により得られるものであることを特徴とする。

【0009】また、請求項4による発明は、請求項1記載の音声信号変形接続方法において、前記互いに異なる声質の音声信号は、一方は人間による発声、もう一方は音声合成装置による発声により得られるものであることを特徴とする。

【0010】また、請求項5による発明は、請求項1記載の音声信号変形接続方法において、前記パラメータは音声のスペクトルであり、前記所定の時間にわたって前記スペクトルを徐々に変形させることを特徴とする。

【0011】また、請求項6による発明は、請求項5記載の音声信号変形接続方法において、前記音声のスペクトルの変形に関しては、前記2つの音声信号間の対応する音素内のピッチの対応を決定し、対応したピッチ毎に、ある周波数を境界周波数として、一方の音声信号のスペクトルにおける前記境界周波数より高域の部分と、他方の音声信号のスペクトルにおける前記境界周波数より低域の部分とを結合させたスペクトルを生成し、生成されたスペクトルを当該ピッチにおけるスペクトルとすると共に、前記境界周波数を単位時間毎に変化させることを特徴とする。

【0012】また、請求項7による発明は、請求項6記載の音声信号変形接続方法において、前記境界周波数の変化は、単位時間毎に一定量増加するように行われることを特徴とする。

【0013】また、請求項8による発明は、請求項6記載の音声信号変形接続方法において、前記境界周波数の変化は、変化開始時の値から変化終了時の値まで徐々に増加するように行われ、前記変化開始時に近い相対的に低域の段階では比較的ゆっくりと、前記変化終了時に近い相対的に高域の段階では比較的早く変化させることを特徴とする。このような変化は人間の聴覚特性によりマッチしており、より自然な声質変化の実現を可能とする。

【0014】また、請求項9による発明は、請求項1記載の音声信号変形接続方法において、前記パラメータは音声の基本周波数であり、前記所定の時間にわたって前記基本周波数を徐々に変化させることを特徴とする。

【0015】また、請求項10による発明は、請求項9記載の音声信号変形接続方法において、前記音声の基本周波数の変化に関しては、前記各音声信号の平均基本周波数を求め、両者の平均基本周波数の差とパラメータを変更する前記所定時間とに基づいて、単位時間当たりに変更すべき基本周波数の値を求め、この値を変化量の単

位として、一方の音声信号の平均基本周波数から他方の音声信号の平均基本周波数へと単位時間毎に変化させることを特徴とする。

【0016】また、請求項11による発明は、請求項1記載の音声信号変形接続方法において、前記パラメータは、音声のスペクトルと音声の基本周波数であり、前記音声のスペクトルの変形に関しては上記請求項6と同様の方法を用い、前記音声の基本周波数の変化に関しては上記請求項10と同様の方法を用いることを特徴とする。

【0017】また、請求項12による発明は、請求項11記載の音声信号変形接続方法において、前記音声のスペクトルと基本周波数とを並行して変化させることを特徴とする。

【0018】

【発明の実施の形態】以下、図面を参照して、本発明の実施形態について説明する。図1(A)～(C)は、互いに声質の異なる2人の話者による音声信号の波形と、これら音声信号を変形接続して得られた音声信号の波形との関係を示す波形図である。本実施形態の処理では、2人の話者(1及び2)に同じテキストを発声させ、これにより得られる音声信号(図1(A)の101、図1(B)の102参照)を接続処理するものとする。

【0019】処理の結果生成された音声信号は、図1(C)の符号103で示されるように、話者1の音声区間と変形接続区間と話者2の音声区間とから構成される。なお、この例では、2人の話者に同じテキストを発声させているが、必ずしも同じテキストを発声させる必要はない。例えば、以下に示す実施形態では、音声の特性を示すパラメータとして、基本周波数とスペクトルとを選び、該2つのパラメータにおける変形を行っているが、スペクトルの変形は行わず、基本周波数だけを変形する場合には、2人の話者が発声するテキストは異なっても構わない。

【0020】図2は、本実施形態による音声信号変形接続方法の全体の処理の流れを示すフローチャートである。2人の話者が発声した音声信号を入力すると、ステップS201では、それぞれの音声信号に音素境界を付与し、ステップS202に進む。ステップS202では、それぞれの音声信号に基本周期を示すピッチマークを付与し、ステップS203に進む。

【0021】ステップS203では、上記ピッチマークに関して、両音声信号の対応した有声音区間に対し、最も近いピッチマーク同士を選ぶことにより、その区間のピッチマーク同士の対応を付ける。この結果、図3に破線で示すように、1対1、1対多、または、多対1の対応づけが得られる。このピッチマークの対応関係は、ピッチ対応テーブル301(図4参照)に記憶される。次に、ステップS204では、対応した音素ごとに音声信号パワーの正規化を行う。以上の処理は、音声収録後に

独立して予め行ってもよいし、変形接続処理の一部として行ってもよい。

【0022】次に、ステップS205では、後述する方法で音声信号の基本周波数の変換を行い、ステップS207に進む。ステップS207では、後述する方法で音声信号のスペクトルの変形を行い、ステップS209に進む。このとき、基本周波数の変更量の設定はステップS206で行われ、スペクトル変形に関する境界周波数の変更量の設定はステップS208で行われる。これらの変更量は時間の関数となっている。

【0023】最後に、ステップS209で、両音声信号が全体的に合成され、合成音を得る。上記ステップS205で使用可能な基本周波数の基本的変換方式としては、様々な方式が提案されているが、その一例としては、文献「E.Moulines, F.Charpentier, "Pitch-Synchronous Waveform Processing Techniques for Text-to-Speech Synthesis using Diphones", Speech Communication, Vol.9, pp.453-467, Dec.1990」で提案されているPSOLA方式がある。

【0024】図4は、上記ステップS207で行われるスペクトル変形処理の一例を示すフローチャートである。同処理において、ステップS302では話者2の音声信号から、また、ステップS303では話者1の音声信号から、上記ステップS203(図2参照)で求められたピッチ対応テーブル301を参照して互いに対応するピッチを選択し、該ピッチ毎に、ピッチ同期信号に同期して音声波形を切り出す。ここでは、話者1から話者2へと徐々に音声を変形しながら接続する場合を例に挙げて説明する。この場合には、ピッチ同期の処理は、話者2のピッチマークの回数だけ行われる。ここで、図3の有声音Zの例に見られるように、話者1の2つのピッチマークが話者2の1つのピッチマークに対応している場合には、話者1の2つのピッチマークのうちどちらか一方を参照して音声波形を切り出す。一方、図3の有声音Yの例に見られるように、話者2の2つのピッチマークが話者1の1つのピッチマークに対応している場合には、話者1の1つのピッチマークにより参照される音声波形を2度切り出す。

【0025】以下、1ピッチ分の処理を説明すると、ステップS304では、ステップS302で切り出した音声波形についてFFTによるスペクトル分析を行う。また、ステップS304と並行して、ステップS305では、ステップS303で切り出した音声波形についてFFTによるスペクトル分析を行う。ステップS306では、ステップS304で求めた話者2のスペクトルのうち、所定の周波数 $\alpha$ Hzより低い帯域の部分を取り出す。ステップS307では、ステップS305で求めた話者1のスペクトルのうち、上記周波数 $\alpha$ Hzより高い帯域の部分を取り出す。

【0026】このステップS306およびS307で取



り出されたスペクトルは、ステップS308において、周波数 $\alpha$  Hzを境界にして結合される。このスペクトルの混合処理は、各FFTで得られたスペクトルの実部と虚部を、それぞれ個別に処理することで行われる。最後に、ステップS309では、両者のスペクトルを混合したスペクトルに対し、IFFTを行い、1ピッチ波形を得る。こうして得られた1ピッチ波形は、上記ステップS209（図2参照）の処理に渡される。

【0027】更に、上記境界周波数を時間的に変化させながら、ピッチ毎に同様のスペクトル混合処理を行うことにより、複数の“1ピッチ波形”が同様にステップS209の処理に渡され、最終的に該ステップS209で音声合成処理される。図5（A）～（C）、図6（A）～（C）、及び図7（A）～（C）は、境界周波数の時間的变化に伴うスペクトル混合の例を示したものである。ここでは、境界周波数 $\alpha$ を図5（B）→図6（B）→図7（B）に示す順序で3段階に変化させたとして、低域側が抽出される話者2の各段階のスペクトルを図5（A）、図6（A）、図7（A）に、高域側が抽出される話者1の各段階のスペクトルを図5（C）、図6（C）、図7（C）に、そして、各段階でのスペクトル結合により得られた混合スペクトルを図5（B）、図6（B）、図7（B）にそれぞれ示す。

【0028】図8は、本実施形態による音声信号変形接続処理において、ステップS205の基本周波数変換処理で変換される平均基本周波数と、ステップS207のスペクトル変形処理において変換される境界周波数との時間変化を示すグラフである。本実施形態では、音声の基本周波数変化の制御としては、話者1と話者2による音声信号の平均基本周波数をそれぞれ求めておき、両平均基本周波数の差とパラメータを変更する所定の時間（変形接続区間）とに基づいて単位時間当たりに変更すべき基本周波数値を求める。そして、該変形接続区間における基本周波数を、図8に示すように、上記2つの平均基本周波数の一方から他方へと、時間的に一定の割合で変化させる。

【0029】また、音声のスペクトルの制御は、境界周波数 $\alpha$  Hzを時間的に一定の割合で変化させる。ここで、平均基本周波数の変更量はステップS206で設定され、また、境界周波数の変更量は、ステップS208で設定される。

【0030】次に、図9（A）～（C）は、図1（A）～（C）に対応した、各音声信号による声紋を示すスペクトログラムである。各スペクトログラムにおいて、横軸は時間（sec）、縦軸は周波数（Hz）、そして、時間と周波数との各交点における濃さ（本紙面上では明確に現れないが）、その時間におけるスペクトルの強さを表す。また、図9（C）に示す合成音声における変形接続区間には、該区間における境界周波数の変化を、符号105で参照されるラインで示した。

【0031】なお、上述のような音声パラメータを変化させる割合は一定である必要はなく、様々な変化パターンが考えられる。例えば、図10に境界周波数 $\alpha$ の変化パターンの別例を示す。この例では、境界周波数が低い段階ではゆっくりと変化させ、境界周波数が高くなるにつれて早く変化させている。人間の聴覚は高域に比べて低域の周波数に重みがかかった特性を有するため、この例のような変化をさせると、人間が聞いた時により一定の変化割合で、すなわち、より自然に声質を変化させることができる。

【0032】以上、この発明の実施形態を図面を参照して詳述してきたが、具体的な構成はこの実施形態に限られるものではなく、この発明の要旨を逸脱しない範囲の設計の変更等があってもこの発明に含まれる。たとえば、上述した一実施形態においては、初めに基本周波数を変化させ、その後に音声スペクトルを変化させているが、変化の順番はこの逆でも構わないし、また、分散処理等により、両者を同時に変化させてもよい。なお、変形区間が長い場合には、これらパラメータの変化を分けて行った方が、より滑らかな接続が行える。

【0033】また、変形接続される2つの音声信号は、人間による発声により得られたもの以外に、音声合成装置による発声で得られたもの同士でもよく、また、人間による発声と音声合成装置による発声で得られた音声信号を接続することも可能である。

【0034】なお、以上の説明において、異なる話者の音声（メッセージ）を接続しても聞き手に違和感を与えない点を強調してきたが、本発明は、声の変化を聞き手に全く意識させないだけでなく、違和感の無い声の変化を聞き手に“意識させる”という利用法もある。例えば、画像処理のモーフィングと呼ばれる処理では、男性の顔の静止画像と女性の顔の静止画像とを用いて、（時間とともに画像を変化させながら）男性の顔を徐々に女性の顔へと変化させることが可能である。このような画像処理技術と本発明による方法とを統合すれば、人間の顔が男性から女性に変化しながら、その声もいつのまにか男性から女性に変化しているというような、見る（聞く）者に不思議な感覚を与えるシミュレーションの実現が可能である。このような技術は、映画やマルチメディア作品等の製作分野において、新しい表現手段として利用できる。

【0035】

【発明の効果】上述のように本発明によれば、音声の特徴量を時間と共に変化させることができる。その結果、話者が異なる2つの音声を接続する場合であっても、接続区間における急激な声質の変化を避けることができ、聞く者にとって違和感なく音声を接続することが可能となる。

【図面の簡単な説明】

【図1】 本発明の一実施形態において、互いに声質の

異なる2つの音声の波形と、これらの音声信号に変形接続処理を施して得られた音声波形との関係を示す波形図である。

【図2】 同実施形態における音声信号変形接続方法の全体の処理の一例を示すフローチャートである。

【図3】 同実施形態において、2つの音声信号間のピッチマークの対応づけを説明するための図である。

【図4】 同実施形態におけるスペクトル変換処理の一例を示すフローチャートである。

【図5】 同実施形態において、ある時間における境界周波数の設定と、該境界周波数における2つのスペクトルの結合を説明するための図である。

【図6】 更に進んだ時間における境界周波数の再設定と、該境界周波数における2つのスペクトルの結合を説明するための図である。

【図7】 更に進んだ時間における境界周波数の再設定と、該境界周波数における2つのスペクトルの結合を説明するための図である。

【図8】 同実施形態による音声信号変形接続処理において、平均基本周波数と境界周波数の時間変化を示すグラフである。

【図9】 図1に示す各音声信号による得られる声紋を示すスペクトログラムである。

【図10】 境界周波数の別の時間変化例を示すグラフである。

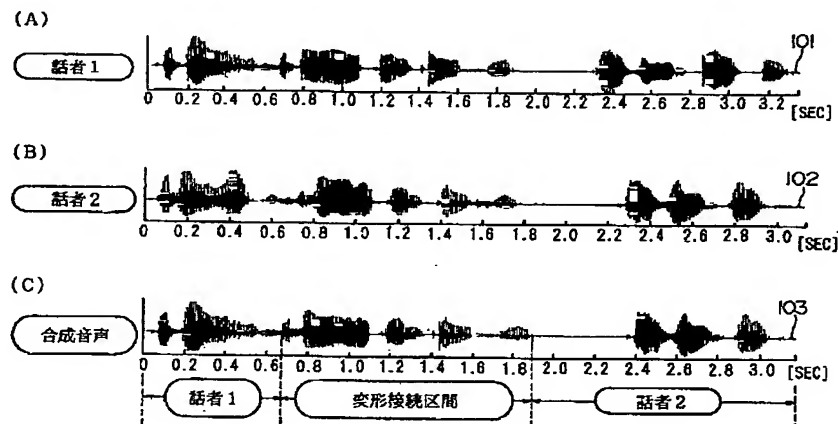
【符号の説明】

101 話者1の音声信号

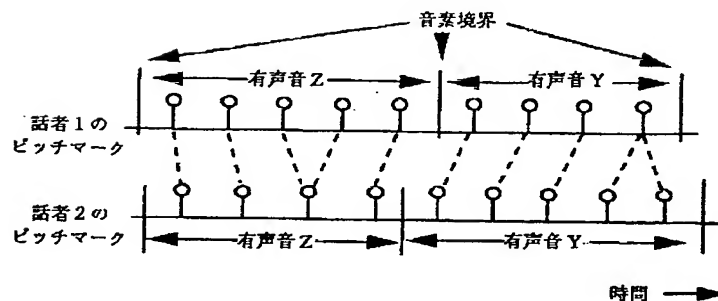
102 話者2の音声信号

103 合成音声信号

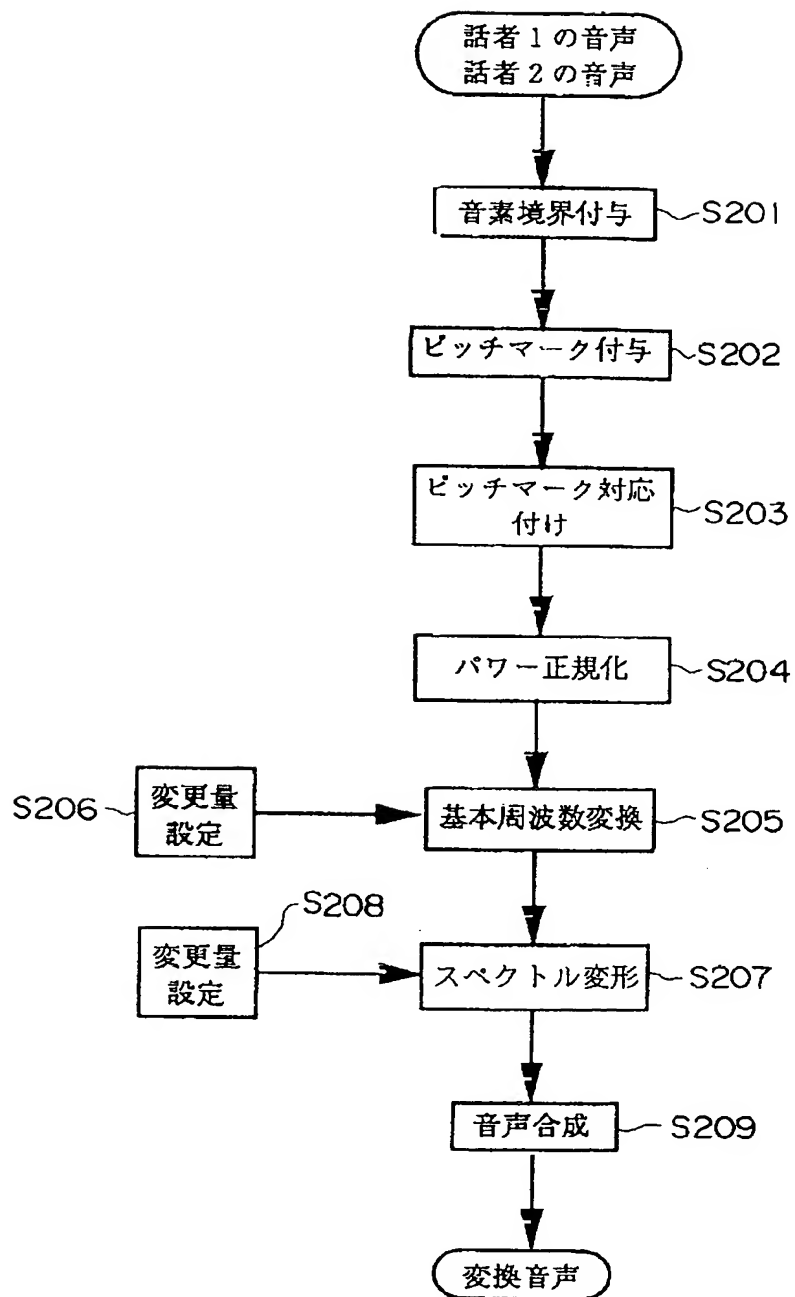
【図1】



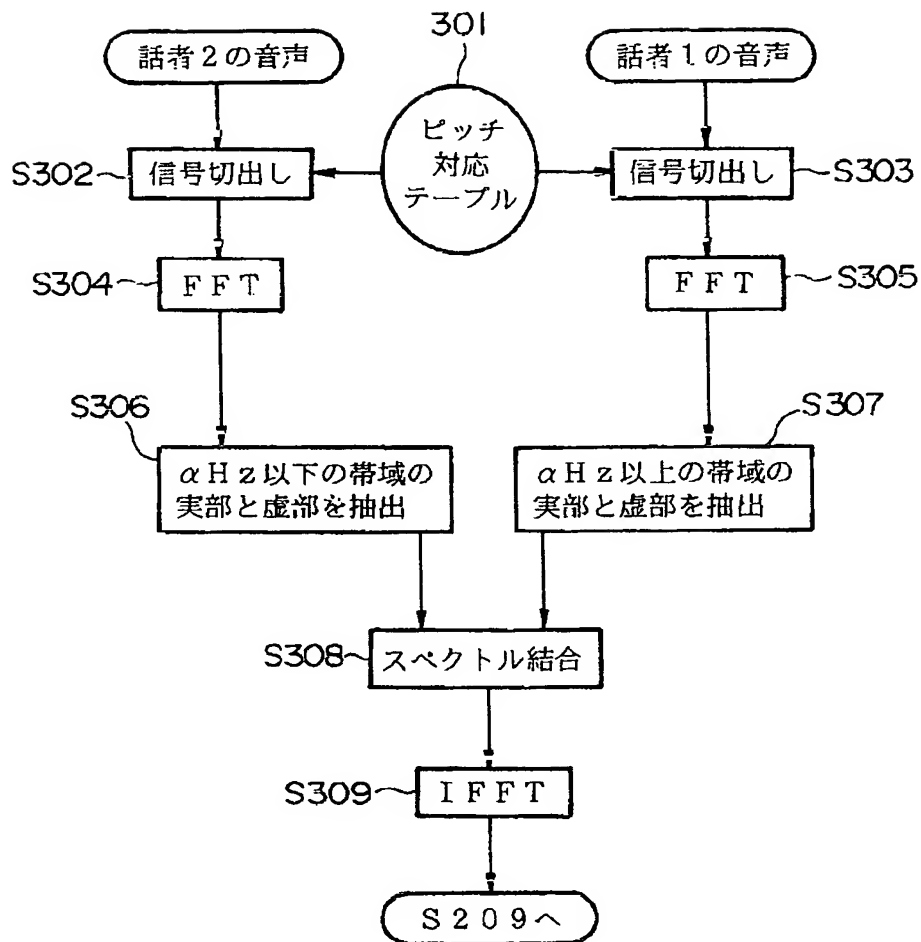
【図3】



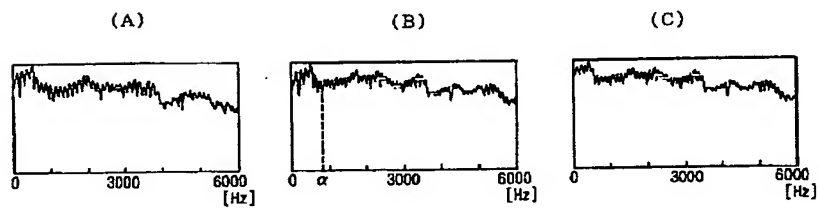
【図2】



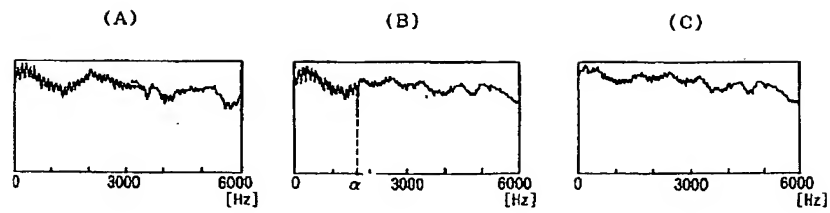
【図4】



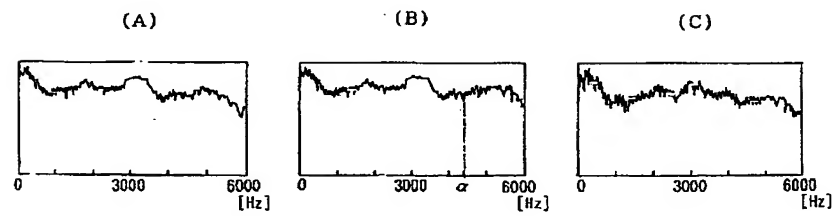
【図5】



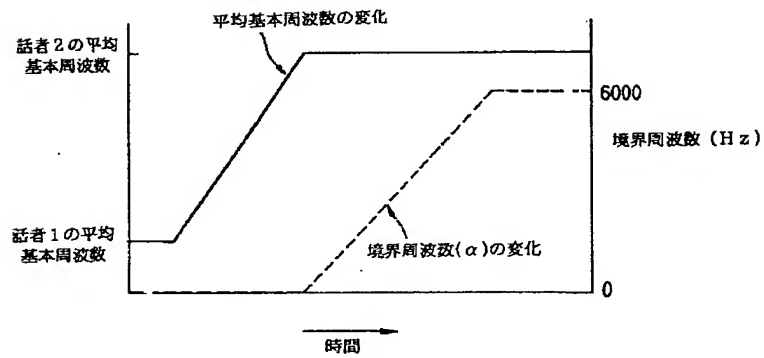
【図6】



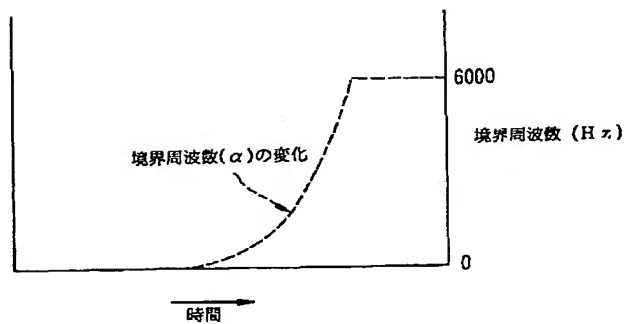
【図7】



【図8】



【図10】



(10)

特開平9-152892

【図9】

